AWARD NUMBER: W81XWH-14-1-0473

TITLE: GENOMIC DIVERSITY AND THE MICROENVIRONMENT AS DRIVERS OF PROGRESSION IN DCIS

PRINCIPAL INVESTIGATOR: EUN-SIL HWANG

CONTRACTING ORGANIZATION: DUKE UNIVERSITY
Durham, NC 27708-4677

REPORT DATE: **October 2017**

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
                        Fort Detrick, Maryland 21702-5012

1. REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2017	Annual	30Sep2016 - 29Sep2017

4. TITLE AND SUBTITLE
Genomic Diversity and the Microenvironment as Drivers of Progression in DCIS

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-14-1-0473

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

E.S. Shelley Hwang

E-Mail: shelley.hwang@dm.duke.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Duke University
2200 W Main St Ste 710
Durham NC 27708-4677

8. PERFORMING ORGANIZATION REPORT

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT The project is designed to test whether genetic and/or tumor environmental heterogeneity is a driving force in progression of breast DCIS. Our project, a collaboration between Duke and ASU, has made substantial progress on all 4 aims and we met our 36 month milestones. Primary achievements for 36 months are:
 1) Continued Case and control identification (45 Pure DCIS & 36 adjacent DCIS with invasion) through extensive database and searching at Duke 2) Deep and comprehensive full exome sequencing for 32 cases from 30-160ng of DNA isolated from archival FFPE specimens, 3) Comparison of analytic methods to characterize somatic mutations from this full exome sequencing, 4) Application of sequencing data for copy number assessment 5) Development of dual immune-staining on DCIS lesions using 7 pairs of antibodies, 6) Imaging analysis of these stains, including quantitative analysis, 7) Identification of upstaged DCIS cases for the radiology aim, 8) Development of image analysis methods for digital mammograms, 9) Validation Aim (4) approval of the Duke IRB/ TBCRC038 protocol at 12 sites, including DOD approval to initiate collection of DCIS that either did or did not progress to invasive cancer, 10) Full integration of team members over the past year via frequent conferencing, face to face meetings, and constant communication. This multi-disciplinary progress puts our group into an ideal position to fully implement the aims of the project and reach our year 4 goals.

15. SUBJECT TERMS
DCIS, intra-tumor heterogeneity, genetic diversity, phenotypic diversity, somatic evolution, microenvironment, mammographic biomarkers

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	21	19b. TELEPHONE NUMBER *(include area code)*
U	U	U			

1.   INTRODUCTION

Ductal carcinoma in situ (DCIS) of the breast is an increasingly common diagnosis that is related to aggressive screening patterns (mammography). This "pre-invasive" lesion may progress to invasive cancer, but does so at a relatively low frequency. Nonetheless, it is commonly treated with extensive surgery, radiation, and hormonal therapy even though most of these lesions would never progress to invasive cancer. Thus, there is a pressing clinical need to stratify the risk of DCIS tumors into those in need of intervention and those that can be safely monitored without intervention. Our project is designed to address this need by characterizing the evolvability of DCIS, detecting those that have a high likelihood of evolving to malignancy versus those that are likely to remain indolent.

2.   KEYWORDS

DCIS, cancer progression, intra-tumor heterogeneity, genetic diversity, phenotypic diversity, somatic evolution, microenvironment, mammographic biomarkers

3.   ACCOMPLISHMENTS

What were the major goals of the project?

Aim 1. *Determine whether genetic diversity of DCIS is greater in DCIS with adjacent invasive disease compared to DCIS without progression.* Diversity measures must be derived from geographically distinct areas of tumor. Genetic divergence of the DCIS component of tumors will be measured based on exome sequencing and SNP arrays run on two separate regions of the tumor, as well as normal tissue, in patients with DCIS either with or without adjacent invasion to determine the association between genetic diversity and progression to malignancy. Genetic diversity will be measured by the genetic divergence between the tumor samples, that is, the proportion of the genome that differs between the two samples from the same tumor.

36 Month Milestones:

- Protocol preparation, IRB submission and approval: **Completed** (Duke eIRB Pro00054515, initial Duke approval, 5/27/2014 and renewed for the current year), DOD IRB approval in place.

- Case identification and tissue block selection: Through a variety of available databases, we identified a large number of cases and controls with tissue available in the Duke Pathology archives. Each potential case and control requires extensive chart and pathology review in order to determine final eligibility and usability. For example, there is sufficient amount of the DCIS lesion (>2mm size) for isolation and DCIS is not too close to invasive cancer (it extends outside the invasive component). There must be two blocks with DCIS present that are >0.8cm apart. To date we have identified **81** cases, pathology reviewed post sections.

- Sectioning of tissue blocks: New sections from candidate paraffin blocks are cut, stained to include one H&E at the beginning and end of each set and then reviewed by the study pathologist. Remaining sections from candidate blocks (containing a sufficient amount of the DCIS lesion of interest) are used for macro-dissection and subsequent DNA extraction. Additional sections (every other one) are also stored for immunohistochemical (IHC) analysis of key measures of tumor and micro-environmental heterogeneity. These slides are scanned for analytic and archival purposes. This process has been fully implemented and we are moving through both cases and controls in this manner.

- DNA extraction of test cases: **Completed**.

- Exome sequencing of test cases: **Completed.** We choose the Genome Center at Washington University who have developed cutting-edge methods for producing high quality data from these FFPE specimens. Over the past two years, **Wash U. sequenced 30-160ng from 153 individual DNA samples derived from 51 subjects** (germ line sample plus 2 DCIS containing samples). They were able to derive interpretable sequence data from 30-160ng of FFPE DNA with qualities summarized in Figure 1, 2 and 3.
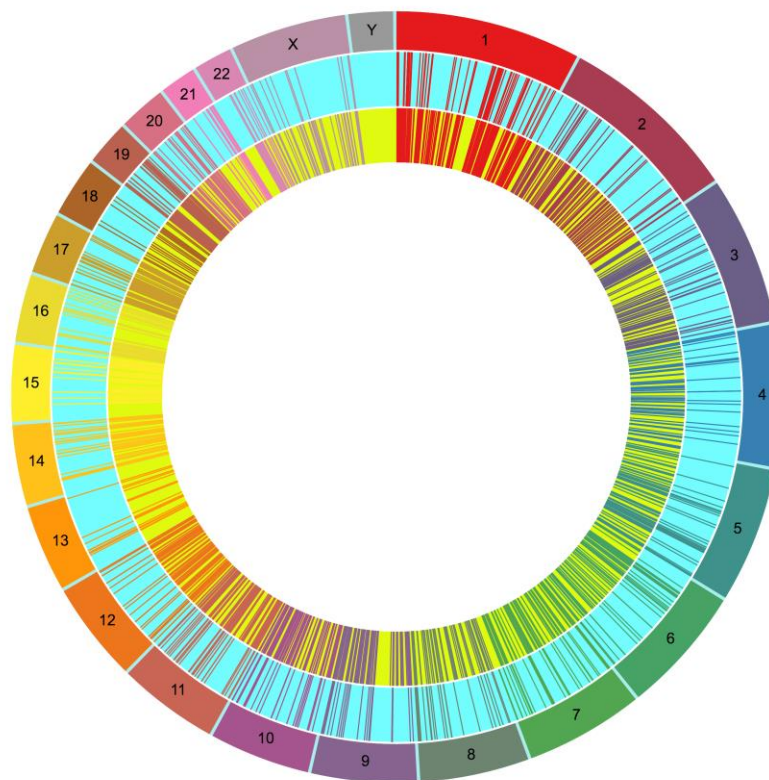

Figure 1: Exomic variants in Outer Track is genome; Middle Track is pure DCIS and Inner Track is adjacent DCIS

- SNP arrays: Since DNA from the primary samples is limiting (macrodissected DCIS), we have been testing whether sequencing libraries generated for exome sequencing can be directly applied to these arrays. We have developed a new pipeline to call copy number variants based on *ASCAT, Copynumber, and Sequenza*. We are now analyzing the results.

- Development of a pipeline for identification of somatic genetic alterations: **Completed.** In order to assess and minimize artefacts induced by the FFPE procedure and the small amounts of DNA obtained from FFPE samples we developed a strategy based on 12 (total of 20 in the pipeline) sequencing technical replicates. Although our pipeline has been completed and is fully functional, we continue to work to improve it. In the last year, these improvements have been statistically significant as seen in Figure 2, Improved SNV Bioinformatics Pipeline (Wilcoxon signed-rank test, $p=0.008$).

Pipeline benchmark using technical replicates

$p=0.008$
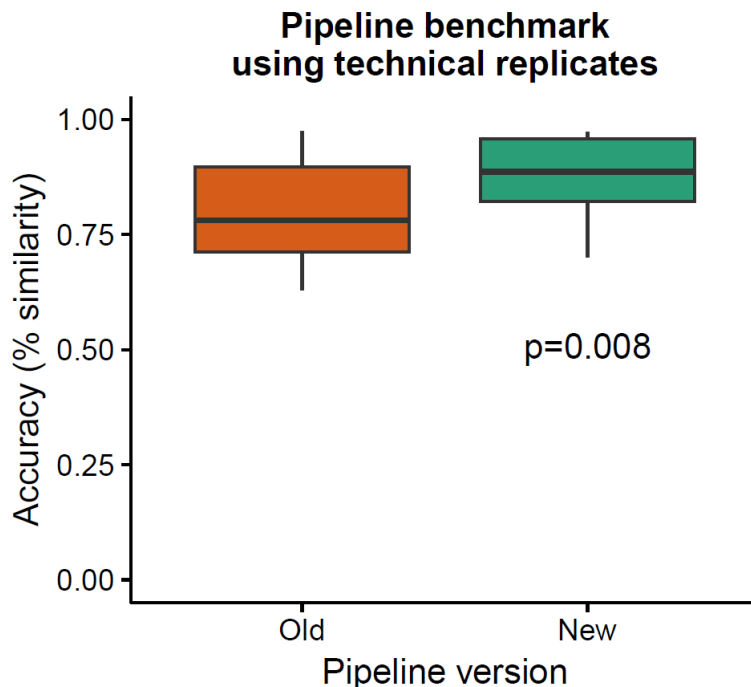
Figure 2: Improved SNV Bioinformatics Pipeline using the Wilcoxon signed-rank test

- Calculation of genetic diversity scores for the pilot cohort: **Completed**. The main purpose of the research project is to determine the heterogeneity between samples. We found a statistically significant higher number of somatic mutations in DCIS adjacent to invasive disease than pure DCIS as seen in Figure 3 (All variants, Welch's t-test, $p=0.027$; exonic variants, $p=0.032$; coding variants, $p=0.031$). We found genes mutated in the majority of patients (e.g, Myomegalin, Trypsin-3) and genes mutated mainly in the DCIS adjacent to

invasive disease (e.g. Dual specificity mitogen-activated protein kinase kinase 3, Leucine-rich repeat serine / threonine-protein kinase 2). Moreover, the DCIS adjacent to invasive disease samples are statistically significant enriched for cell-matrix /cell-cell adhesion biological processes and pathways. Current analysis of genetic heterogeneity suggests that the genetic variability in DCIS adjacent samples was accumulated in the early phase of cancer development and then maintained during the subsequent tumor expansion.
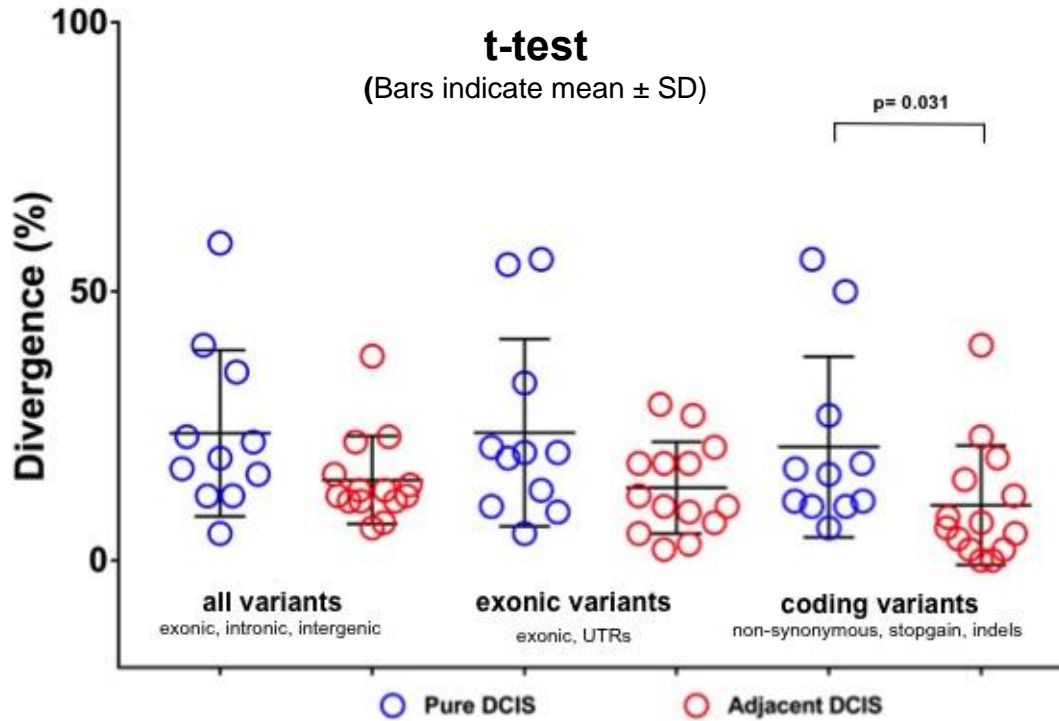

Figure 3: Exomic variants in Pure DCIS vs. adjacent DCIS to invasive disease (t-test)

Aim 2. Determine whether phenotypic diversity of DCIS and the tumor microenvironment (TME) is greater in DCIS with adjacent IDC compared to DCIS without IDC.

Since genomics is not the sole driver of tumor behavior, we will phenotypically characterize DCIS and its microenvironment including markers of hypoxia, migration, proliferation, matrix organization, and immune signaling in the same samples used in Aim 1. We will employ automated image analysis to compute microenvironmental divergence to determine if specific components of the TME, or the divergence between TMEs from the same tumor, differs between DCIS with and DCIS without adjacent IDC.

In the past 12 months, we have analyzed our phenotypic diversity markers to 46 cases with another 8 cases in progress (Table 1). These markers, now including nuclear grade (essentially nuclear size of the DCIS epithelial cells) are shown in Table 2 below. These elements include the presence and distribution of cell types (malignant epithelia, lymphocytes, and stroma) and expression of proteins that are associated with oncogenic and environmental processes. To

evaluate these elements, we are using a combination of expert scoring and automated image analysis. Further, expert scoring is being used to guide, train, and evaluate the image analysis so these analyses will be cross-informative.

Table 1: Cohort Demographics

	Pure DCIS	Adjacent DCIS	
Age (mean)	56.1	56.9	
Race			
White	17	16	
Black	4	2	
Other	1	2	
Tumor Size (mean, cm)	4.3	4.2	
Nodal Status			
Negative	22	11	
Positive	0	9	
Grade		*Invasive*	*DCIS*
1	0	3	0
2	11	6	7
3	11	11	13
Surgery			
Lumpectomy	14	9	
Mastectomy	8	11	
Estrogen Receptor			
Positive	15	16	
Negative	4	3	
Equivocal	3	1	
HER2			
Positive	4	4	
Negative	14	10	
Equivocal	4	6	

Table 2: Phenotypic Markers of Heterogeneity & Pathology Scoring

Stain	Marker	Function	Cell Type	Scoring
Double				
	ALDH1A1	Stem Cell Marker	Epithelia	Intensity + Distribution
	Ki-67	Proliferation	Epithelia	Distribution
	COL15A1	Basement Membrane	BM	Presence around DCIS
	ESR1	Hormone Signaling	Epithelia	Intensity + Distribution
	Phospho-FAK	Cell Adhesion	Epithelia	Intensity + Distribution
	CD68	Macrophage	Macrophage	Distribution
	CA9	Hypoxia	Epithelia	Intensity + Distribution
	FOXP3	T Regulatory Cells	Lymphocyte	Distribution
	ERBB2	Oncogenic Signaling	Epithelia	Intensity + Distribution
	P63	Basal Cells	Myoepithelia	Presence around DCIS
	RANK	Inflammatory Signaling	Epithelia	Intensity + Distribution
	PGR	Hormone Signaling	Epithelia	Intensity + Distribution
Single				
	GLUT1	Glucose Transport	Epithelia	Intensity + Distribution
	CD31	Blood Vessels	Endothelia	Distribution
	Rho A	Motility	Epithelia	Intensity + Distribution
Other				
	Nuclear Grade	Histologic Categorization	Epithelia	Distribution

The image analysis pipeline for cell type identification and enumeration is now set and incorporates the following elements: CRimage package implemented in R, cell segmentation based on the watershed algorithm, 150 morphology and texture features computed for each cell, and a

support vector machine classifier that is trained using cell identification from our study pathologist (Dr. Allison Hall). Based on 8307 cell-level hand annotations by our pathologist, we have trained an automated image analysis framework capable of segmenting and differentiating DCIS epithelial cells, lymphocytes, and stromal fibroblasts with accuracy of 90.4% on an independent test set, consisting of 3033 cells from four samples (Figure 4). A three cell type classifier is also being applied (without endothelial cells) and this reaches an accuracy of over 90%. The pipeline is able to accommodate variations in H&E staining by utilizing a stain normalization method and masking artefacts from ink or blood. We are currently evaluating a number of spatial statistics methods to identify microenvironmental features discriminative of regions of pure DCIS and adjacent DCIS to invasive disease.
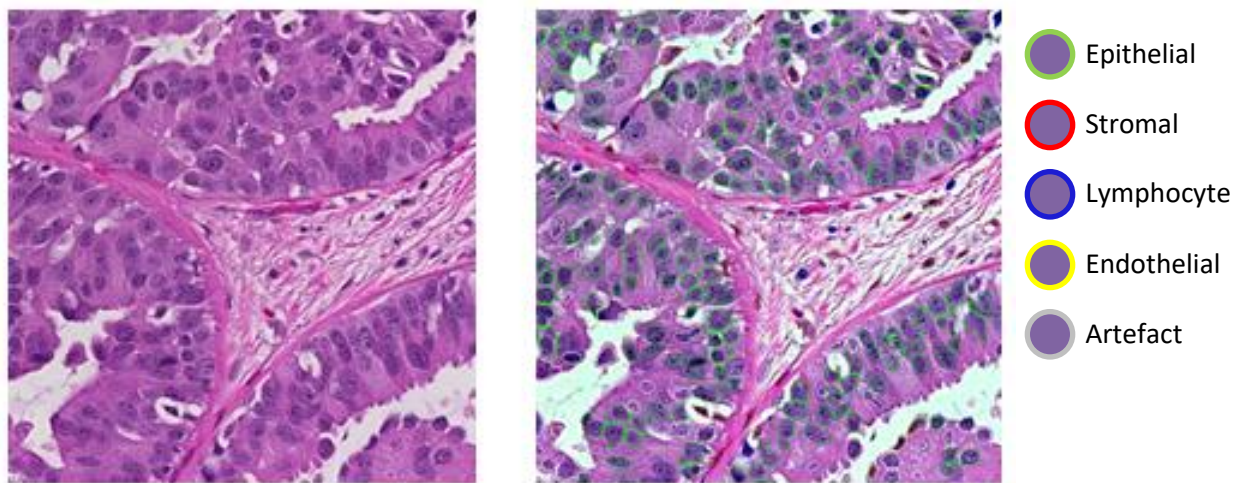

Figure 4. Original image and cell classification results. Green indicate epithelial cells, blue - lymphocytes, red - stromal cells, yellow – endothelial cells, and white - segmentation artefacts.

The image analysis pipeline for IHC analysis incorporates the following elements: a deep learning algorithm creates a mask to produce a mask for tumor regions, a modified version of CRimage is used to classify the cells (see above), IHC staining is quantified in two channels (brown for nuclear and single stains, red for the non-nuclear stains).

Thirty four cases have been evaluated by expert scoring of our study pathologist. This is an interim analysis of our planned 100 cases so results must be interpreted with caution Table 3. We evaluated these data in two ways: 1) Is there an overall difference, in the DCIS component, between pure DCIS and adjacent DCIS and 2) Is there evidence of difference in heterogeneity of these phenotypic markers between pure and adjacent DCIS. Distributional heterogeneity was measured using the earth mover difference (EMD) test.

Some of the interesting results from these interim analyses are shown below. Notable findings include: 1) More intact myoepithelial layer (p63 staining) and constitution of the basement membrane (COL15A) in the pure DCIS, 2) Higher proliferation in the adjacent DCIS, 3) Higher levels of HER2 and PR in pure DCIS, and 4) Increased distributional heterogeneity in nuclear grade in adjacent DCIS. Each of these may be useful markers for discriminating DCIS likely to progress which will be tested in Aim 4.

Table 3: Histologic Parameters

Parameter	Overall Comparison[3]			Heterogeneity (EMD)[4]		
	Pure	Adjacent	p-Value	Pure	Adjacent	p-Value
Nuclear Grade				0.13	0.29	**0.05**
KI67 (% positive)	12.3	19.3	**0.03**	0.37	0.56	0.24
COL15A1 (duct ringing)	0.25	0.05	**0.002**	0.29	0.07	**0.02**
TP63 (myoepithelial)	0.42	0.31	**0.01**	0.44	0.43	0.97
Estrogen Receptor [1]	127	95	0.23	0.3	0.27	0.86
Progesterone Receptor[1]	96	54	**0.04**	0.47	0.29	0.28
HER2[1]	90	29	**0.01**	0.16	0.13	0.76
GLUT1[1]	111	120	0.69	0.3	0.46	0.09
CD68[2]	7.5	9.7	0.15	0.56	0.73	0.51
FOXP3[2]	6.1	5.9	0.83	0.47	0.31	0.52
CD31[2]	21	22	0.92	0.52	0.57	0.85

[1]H Score
[2]Average or maximum number/HPF
[3]Average values of the two classes, p values from t-tests
[4]Earth mover distance (EMD), p values from t-tests between the two classes

36 Month Milestones:

- IHC staining of candidate markers on all cases: We have obtained and characterized a series of antibodies representing our initial targets including ER, PR, KI-67, COL15A1, RHOA, RAC, CA9, HIF1a, FOXP3, and cleaved Caspase 3. We have piloted dual staining for sets of these antibodies on test cases of breast cancer and will soon be staining for these antigens on DCIS cases and controls. Dual staining conditions will be optimized in collaboration with Dr. Yinyin Yuan's lab who will perform the automated, quantitative scoring and analysis of the stained tissues.

- Scan IHC and H&E stained slides for Automated image analysis (AIA)

- Training and validation of AIA for the identification and enumeration of cell types (epithelial, stromal, lymphocytes, blood vessels). Computer algorithms are trained by expert identification of cell types (study pathologist, Allison Hall). Accuracy of the computer identification is evaluated by comparison back to the expert scoring. To date, we have achieved accuracies of over 80% for this challenging application.

- Develop methods for agnostic computer scoring of IHC stains. These methods are now in testing phase by Dr. Yuan's post-doctoral fellow, Violet Kovacheva and will be implemented on all images.

- Develop computer vision methods to measure nuclear size of the epithelial component. These methods have been developed by Dr. Yuan's team and are in testing phase. All cases will be analyzed for this parameter by Dr. Kovacheva.

Aim 3. Create and test a computational learning algorithm to compare mammographic characteristics and diversity measures in pure DCIS compared to DCIS with IDC. A weighted computational algorithm using mammographic features of lesional and stromal characteristics as well as heterogeneity measures derived from Aims 1 and 2 will be constructed. The tool will be designed to allow for radiologic discrimination between good and poor prognosis DCIS, and will be evaluated in a validation set.

36 Month Milestones:

- We published the first journal paper from this aim (Shi *et al.*, Academic Radiology 2017, PMC5557686). In that study, we extracted 113 conventional, computer vision features and then used a logistic regression model to predict pure DCIS (negative) vs. DCIS upstaged at definitive surgery to reveal occult invasion (positive). This model performed with receiver operating characteristic (ROC) area under the curve (AUC) of 0.70 (Figure 5). These conventional features were designed and selected using a laborious, "handcrafted" approach that is typical of conventional statistical approaches. This intentionally conservative approach provides a baseline for comparing against subsequent studies.
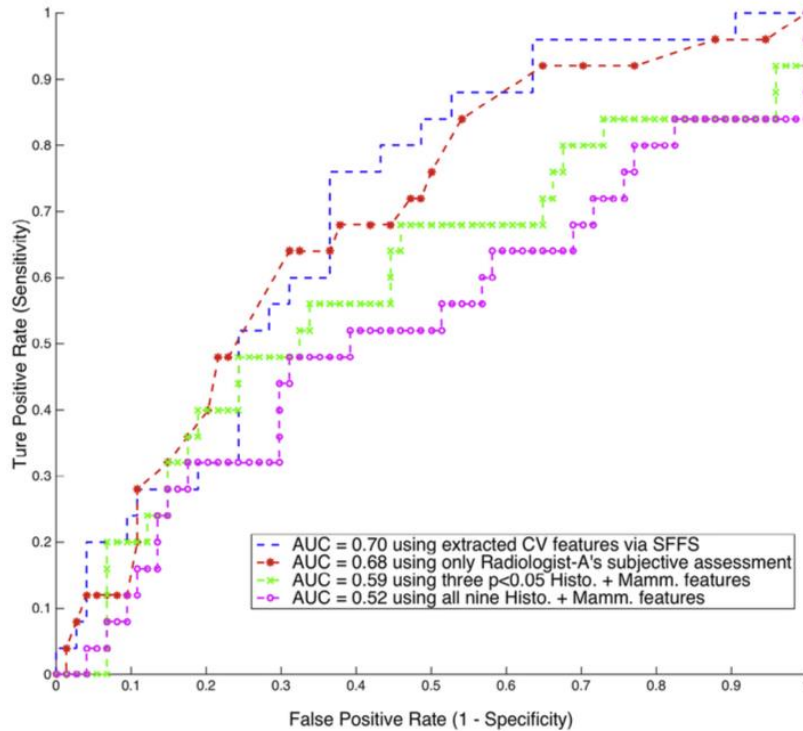
Figure 5. ROC curves for computer vision features. The best performance (AUC=0.70) was for a "handcrafted" subset of computer vision features.

- To our previous cohort of 99 cases, we added 41 cases to increase our data set to 105 pure DCIS and 35 upstaged for a total of 140 cases. This data set is now being used for training and cross-validation for all studies described below. We will reserve the remaining 100 cases for model testing.

- We conducted a new study to perform the classification task using deep learning features. Unlike conventional deep learning approaches that require massive numbers of cases, which are not available for this task, we instead investigated "transfer learning" of the knowledge contained within existing models that have been extensively optimized for unrelated tasks, e.g., natural object classification. These pre-trained convolutional neural networks are comprised of many layers, in which the filters focus progressively on edges, textures, and finally patterns. By feeding our DCIS images into the network, we can use the intermediate filter responses as new features. We hypothesize that these complex filters may be able to characterize subtle patterns of tumor heterogeneity, and to do so better than conventional, handcrafted features.

- Based on initial results from the transfer learning study, we submitted a second paper, which has been accepted for a special issue focusing on deep learning in medical imaging: Shi *et al.,* J Am Coll Radiol 2017. In this study, we pre-trained the deep model for three different classification tasks that are increasingly more similar to our task: ImageNet natural images, Describable Textures Database (DTD) textures, and INbreast digital mammography BI-RADS assessments. After feeding in our DCIS images, the deep

features were synthesized using a logistic regression classifier as before. We hypothesized that more similar tasks in pre-training would lead to better performance for our tasks as well, and this was supported by our results. In the order of increasing relevance, the AUC results for predicting pure DCIS vs. upstaging were ImageNet 0.70, DTD 0.73, and IN breast 0.74 (Figure 6). Note that these results all match or exceed that of our previous, baseline study.
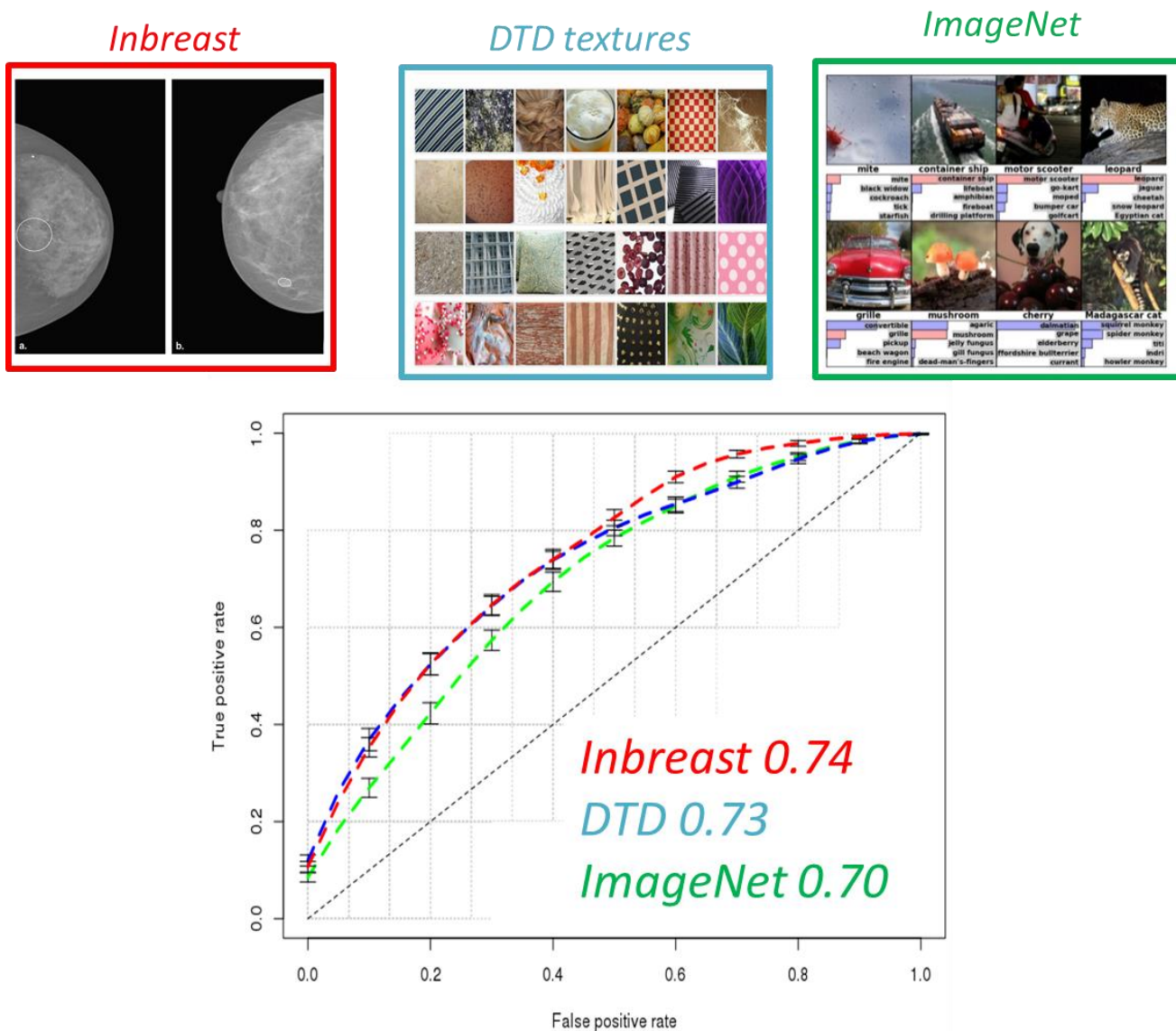

Figure 6. Image databases used for pre-training a deep learning model on unrelated tasks: (left to right) IN breast mammography assessments, DTD textures and ImageNet natural images,. Features from deep layers were subsequently used for our own DCIS classification task.

- We initiated a third study using "forced labeling" of neighboring classes. Given the difficulty of classifying pure DCIS (negative) vs. upstaged DCIS (positive), we added cases of ADH as "super-negative" and IDC as "super-positive" cases, re-labeling them as negative and positive cases, respectively. We hypothesize there is a relationship in image

appearance across these 4 classes, and that the more obvious extremes of ADH and IDC can inform the differentiation of the more subtle pure DCIS vs. upstaged cases (Figure 7). Preliminary studies show that adding IDC cases alone do not improve performance, adding ADH alone provides marginal improvement, but adding both together provides the greatest improvement.
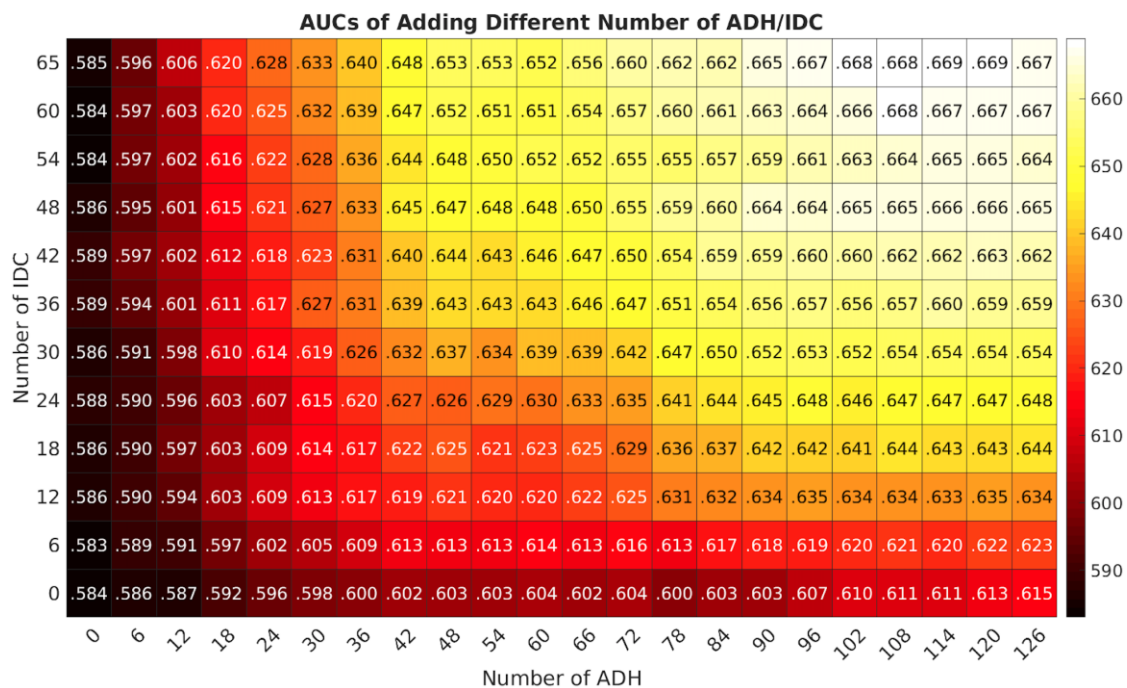

Figure 7. For our task of predicting DCIS vs. upstaged, heat map shows results of adding different numbers of cases of ADH (super-negatives) on the x-axis and IDC (super-positives) on y-axis. Lower left corner represents baseline without any added cases, greatest improvements come from addition of both at upper right corner.

Aim 4. *Test the predictive performance of the best diversity measures in an independent validation set of pure DCIS with and without subsequent invasive recurrence.* Genotypic and phenotypic measures of diversity derived from Aims 1-2 will be applied to an independent case-control, longitudinal, tissue bank of DCIS with and without invasive recurrence to validate their utility. The Duke IRB approved protocol has been approved at 12 sites. For the next budget year, we will continue to accrue cases of pure DCIS that are long term disease free or recurred with invasive cancer. Slides are being shipped to Duke for macrodissection for DNA analysis and for immunodetection of phenotypic heterogeneity.

36 Month Milestones: This aim will be carried out after aims 1-3 are complete. We obtained approval to obtain these specimens through the Translational Breast Cancer Research Consortium (TBCRC) and Duke IRB approval. We have identified 12 high volume academic medical center members of the consortium who obtained regulatory approval, DOD approval and completed an SIV training.

We have finalized a REDCap database for data entry online and slide inventory control. The REDCap online data entry forms are operational, MTA's are in place in 11 of the 12 sites and contracts are executed in all 12 participating sites. We started accrual of outside cases last July. Slides are being shipped to Duke for macrodissection, then DNA analysis and immunodetection of phenotypic heterogeneity. Overall, this aspect of the project is adhering to our proposed timeline and should achieve its accrual and analysis goals.

Below is the list of centers that have agreed to participate in this study in addition to Duke:

Table 4: Multicenter Site Update

Site Name	PI	IRB	MTA	CONTRACT	SIV	DOD Approval
Baylor	Nangia, Julie	Approved	Executed	Executed	Complete	Approved
Chicago	Rita Nanda	Approved	Executed	Executed	Complete	Approved
DFCI	Tari King, MD	Approved	Executed	Executed	Complete	Approved
Georgetown	Shawna Willey	Approved	Executed	Executed	Complete	Approved
Indiana	Anna Maria Storniolo	Approved	Executed	Executed	Complete	Approved
Mayo	Fergus Couch	Approved	pending	Executed	Complete	Approved
MDACC	Alastair M. Thompson	Approved	Executed	Executed	Complete	Approved
Montefiore	Bryan Harmon	Approved	Executed	Executed	Complete	Approved
Pittsburgh	Priscilla McAuliffe	Approved	Executed	Executed	Complete	Approved
UNC	Kristalyn Gallagher	Approved	Executed	Executed	Complete	Approved
UWashington	Dr. Mark Kilgore	Approved	Executed	Executed	Complete	Approved
UPENN	Angela DeMichele	Approved	Executed	Executed	Complete	Approved

What was accomplished under these goals?

Our primary goals have been met including, most importantly, identifying the most efficient method of sequence generation from small amounts of fixed DNA. We have acquired more radiology imaging data sets and established the computer vision algorithms for their analysis. Further, based on our databases, we are confident of accruing sufficient cases and controls at

Duke to fulfill the Aim 1 and 2 goals of the project. Overall, we are in excellent position to complete the proposed work in the project period along the time line that was provided.

What opportunities for training and professional development has the project provided?

We hired several new post-doctoral fellows in the previous year to continue expanding our analysis. Violet Kovacheva has acquired new skills in deep learning methods and attended a conference on breast cancer diagnosis. Bibo Shi has acquired new skills in medical image analysis and continues to learn about the complexities of breast cancer diagnosis.

How were the results disseminated to communities of interest?

We had two DCIS abstracts based on aims 1 and 2 presented at the San Antonio Breast Cancer Symposium in December 2016.

Bibo Shi presented 2 talks at the SPIE Medical Imaging 2017 conference, was accepted for upcoming talk at SABCS 2017 and poster at SPIE Medical Imaging 2018. Rui Hou was accepted for a talk at SPIE Medical Imaging 2018.

What do you plan to do during the next reporting period to accomplish the goals?

Aim 1: We will continue to identify potential cases and controls through Duke Pathology archives and databases and carefully examine each subject for their eligibility. Diagnostic slides from candidate subjects will be evaluated by our study pathologist to determine if there is sufficient material to work with and ones that pass this metric will be included in the study. New unstained slides will be ordered from these cases for macrodissection and immunohistochemical staining. DNA extracted from these slides will be exome sequenced and applied to SNP arrays. Returned data from these assays will be analyzed using our current pipeline in order to scale up from the pilot study to a study with a bigger sample size, which will allow us to get more insights from the data. Moreover, we will investigate the biological meaning of the most common variants of the two different tumor types. We will also continue to improve our sequencing analysis pipeline by analyzing additional technical replicates. We will describe this novel method of analysis of genomic sequence from small amounts of DNA extracted from FFPE samples in a manuscript.

Aim 2:
We will complete the dual IHC staining on the remaining cases, as they come off line after pathology review. Improve methods for agnostic computer scoring of IHC stains. These methods will be implemented on all images. Develop computer vision methods to measure nuclear size of the epithelial component. These methods have been developed by Dr. Yuan's team and are in testing phase. All cases will be analyzed for this parameter by Dr. Kovacheva.

Aim 3:
We will submit another paper describing the final results of the transfer learning of deep features. We will complete the analysis of the forced labeling study to improve classification by addition of neighboring classes, and submit that as an additional paper. We will then perform the majority of the final modeling studies using all cases from our institution, as well as begin to analyze cases from other institutions.

Aim 4:
This multicenter validation arm of the project is set up through the Translational Breast Cancer Research Consortium (TBCRC), a collaborative group set up to conduct innovative and high-impact breast cancer clinical trials.
The validation protocol has been approved by both the TBCRC and the Duke IRB (3/18/2016). Twelve (13 including Duke) external sites have obtained local IRB approval. Sites have both IRB as well as DOD approval and completed an SIV call training session with key personnel from each site,
We will continue to collect cases from sites, where each site will supply approximately 10-12 cases and controls including unstained sections from two DCIS blocks and one germ line block. We have 18 candidate cases at Duke that were pathology confirmed post sectioning. We have 36 cases in the RedCap database from other sites. We currently participate in monthly calls with TBCRC participating sites (17) where clinical coordinators, from all active TBCRC studies, provide updates and questions are addressed.

4.   IMPACT
Successful completion of this project will lead to a variety of biomarkers (genetic, IHC and radiographic) to distinguish high risk from low risk DCIS. This would reduce patient suffering and conserve clinical resources for the women with low risk DCIS, and focus management efforts and clinical resources on women with high risk disease, potentially justifying the risks of interventions. As the project is in its initial stages, these important impacts await in the future.

What was the impact on the development of the principal discipline(s) of the project?

Nothing to report.

What was the impact on other disciplines?

Nothing to report.

What was the impact on technology transfer?

Nothing to report.

What was the impact on society beyond science and technology?

Nothing to report.

5.   CHANGES/PROBLEMS

Changes in approach and reasons for change

There have been no changes in approach.

Actual or anticipated problems or delays and actions or plans to resolve them

So far the problems that have emerged have been primarily technical. Full exome sequencing from small amounts of FFPE tissue is at the limit of current technical practice. Wash U. was the only facility able to do this, of the ones that we tested. That worked well initially, but when Elaine Mardis left Wash. U. their methods suffered and we went through several months in which we could not get reliable data from them. Since then, they have identified and corrected the problems, to the point that we are getting even better results than we did initially. The result was fewer samples processed in the past year, but we are now back on track to complete the proposed work under or original timeline.

Changes that had a significant impact on expenditures

None

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

None

Significant changes in use or care of human subjects

None

Significant changes in use or care of vertebrate animals.

Not applicable.

Significant changes in use of biohazards and/or select agents

Not applicable

6.   PRODUCTS

Publications

1. Walther, V., Hiley, C.T., Shibata, D., Swanton, C., Turner, P.E., and **Maley, C.C.**: Can oncology recapitulate paleontology? Lessons from species extinctions. Nature Reviews Clinical Oncology, 12:273-285, 2015. doi:10.1038/nrclinonc.2015.12. Published. Acknowledged federal support.

2. Caulin, A.F., **Maley, C.C.:** Solutions to Peto's Paradox Revealed by Mathematical Modeling and Cross-Species Cancer Gene Analysis. Philosophical Transactions of the Royal Society of London B, 370 (1673):20140222. Published. Acknowledged federal support.

3. Aktipis, C.A., Boddy, A.M., Jansen, G., Hibner, U., Hochberg, M.E., **Maley, C.C.**, Wilkinson, G.S.: Cancer across the tree of life: Cooperation and cheating in multicellularity. Philosophical Transactions of the Royal Society of London B, 370 (1673):20140219. Published. Acknowledged federal support.

4. Noemi Andor, Trevor A. Graham, Marnix Jansen, Li C. Xia, C. Athena Aktipis, Claudia Petritsch, Hanlee P. Ji, **Carlo C. Maley**: Pan-cancer analysis of the extent and consequences of intra-tumor heterogeneity. Published. Nature Medicine 22:105-13, 2016. Acknowledged federal support.

5. **Carlo C. Maley**, Konrad Koelble, Rachael Natrajan, Athena Aktipis and Yinyin Yuan: An ecological measure of immune-cancer colocalization as a prognostic factor for breast cancer. Breast Cancer Research 17:1-13, 2015. Published. Acknowledged federal support.

6. Shi B, Grimm LJ, Mazurowski MA, Baker JA, Marks JR, King LM, Maley CC, Hwang ES, **Lo JY**, "Can Occult Invasive Disease in Ductal Carcinoma In Situ Be Predicted Using Computer-extracted Mammographic Features?" Academic Radiology, 24 (9), 1139-1147 (2017). PMC5557686. Published. Acknowledged federal support.

7. Shi B, Grimm LJ, Mazurowski MA, Baker JA, Marks JR, King LM, Maley CC, Hwang ES, **Lo JY**, Prediction of occult invasive disease in ductal carcinoma in situ using deep learning features, J Am Coll Radiol, accepted (2017). Acknowledged federal support

8. Shi B, Grimm LJ, Mazurowski MA, Marks JR, King LM, Maley CC, Hwang ES, **Lo JY**, Prediction of occult invasive disease in ductal carcinoma in situ using computer-extracted mammographic features, Proc. SPIE 10134, Medical Imaging 2017: Computer-Aided Diagnosis, Armato SG, Petrick NA, Eds., 101341I (2017). Published. Acknowledged federal support.

9. Shi B, Grimm LJ, Mazurowski MA, Marks JR, King LM, Maley CC, Hwang ES, **Lo JY**, "Can upstaging of ductal carcinoma in situ be predicted at biopsy by histologic and mammographic features?" Proc. SPIE 10134, Medical Imaging 2017: Computer-Aided Diagnosis, Armato SG, Petrick NA, Eds., 101342X (2017). Published. Acknowledged federal support.

10. Abegglen, L.M., Caulin, A.F., Chan, A., Lee, K., Robinson, R., Campbell, M.S., Kiso, W.K., Schmitt, D.L., Waddell, P.J., Bhaskara, S., Jensen, S.T., **Maley, C.C.**†, Schiffman , J. D.†: Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. JAMA, 314:1850-1860,

2015. Published. Acknowledged federal support.

11. Li, X., Paulson, T.P., Galipeau, P.C., Sanchez, C.A., Liu, K., Kuhner, M.K., **Maley, C.C**., Self, S.G., Vaughan, T.L., Reid, B.J., Blount, P.L.,: Assessment of esophageal adenocarcinoma risk using somatic chromosome alterations in longitudinal samples in Barrett's esophagus. Cancer Prevention Research, 8:845-56, 2015. Published. Acknowledged federal support.

12. Kostadinov, R., **Maley, C.C**., Kuhner, M.K.: Bulk genotyping of biopsies can create spurious evidence for hetereogeneity in mutation content. PLoS Computational Biology, 12:e1004413, 2016. Published. Acknowledged federal support.

13. Andor, N., **Maley, C.C**., Ji, H. P. Genomic Instability in Cancer: Teetering on the Limit of Tolerance. Cancer Research 77:2179-2185, 2017. Published. Acknowledged federal support.

14. Tollis, M., Boddy, A. M., **Maley, C.C ,**. Peto's Paradox: How has evolution solved the problem of cancer prevention? BMC Biology 15:60, 2017. Published. Acknowledged federal support.

15. Fortunato, A., Boddy, A., Mallo, D., Aktipis, A., **Maley, C.C**. †, & Pepper, J. W. †: Natural Selection in Cancer Biology: From Molecular Snowflakes to Trait Hallmarks. Cold Spring Harbor Perspectives in Medicine, a029652, 2016. († = co-senior authors) Published. Acknowledged federal support.

16. **Maley, C.C**., Aktipis, A., Graham, T.A., Sottoriva, A., Boddy, A.M., Janiszewska, M., Silva, A.S., Gerlinger, M., Yuan, Y., Pienta, K.J., Anderson, K.S., Gatenby, R., Swanton, C., Posada, D., Wu, C.-I., Schiffman, J.D., Hwang, E.S., Polyak, K., Anderson, A.R.A., Brown, J.S., Greaves, M., Shibata, D.: Classifying the Evolutionary and Ecological Features of Neoplasms. Nature Reviews Cancer, Sept. 15, 2017. Published. Acknowledged federal support.

Technologies or techniques

Nothing to report

Inventions, patent applications, and/or licenses

Nothing to report

Other Products

Case report forms for Duke and outside cases and databases to efficiently capture this information

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Co-PI: Dr. Shelley Hwang (M.D., M.P.H.): Duke University (no change)
Co-PI: Dr. Carlo C. Maley (Ph.D.): Arizona State University (no change)

Co-Investigators:
Dr. Jeffrey Marks (Ph.D.): Duke University (no change)
Dr. Joseph Geradts (M.D.): Duke University (departed during year one)
Dr. Allison Hall (M.D.): Duke University, replacing Dr. Geradts.
Dr. Joseph Lo (Ph.D.): Duke University (no change)
Dr. Jay Baker (M.D.): Duke University (no change)
Dr. Yin Yin Yuan (PhD): Institute for Cancer Research, UK (no change)
Dr. Lars Grimm (M.D.): Duke University (no change)
Dr. Trevor Graham (Ph.D.): Barts Cancer Institute, Queen Mary University of London (no change)
Dr. C. Athena Aktipis (Ph.D.): Arizona State University (no change)
Dr. Shane Jensen (Ph.D.): University of Pennsylvania (departed during year one)

Post-Docs:
Dr. Mengyu Wang (PhD): Duke University (departed during year one)
Dr. Violet Kovacheva (PhD): Institute for Cancer Research, UK (no change)
Dr. Lorraine King (PhD): Duke University (no change)
Dr. Bibo Shi (PhD): Duke University (no change)
Rui Hou, ECE Ph.D. student, Duke University

Dr. Angelo Fortunato (PhD): Arizona State University (no change)
Dr. Diego Mallo (PhD): Arizona State University (no change)